DOI:
10.1093/ijlct/ctv023

Document Version
Publisher's PDF, also known as Version of record

Design and optimization of a thermoacoustic heat engine using reinforcement learning

Jurriath-Azmathi Mumith, Tassos Karayiannis and Charalampos Makatsoris*

Department of Mechanical, Aerospace and Civil Engineering, College of Engineering, Design and Physical Sciences, Brunel University London, London, UK

Abstract

The thermoacoustic heat engine (TAHE) is a type of prime mover that converts thermal power to acoustic power. It is composed of two heat exchangers (the devices heat source and sink), some kind of porous medium where the conversion of power takes place and a tube that houses the acoustic wave produced. Its simple design and the fact that it is one of a few prime movers that do not require moving parts make such a device an attractive alternative for many practical applications. The acoustic power produced by the TAHE can be used to generate electricity, drive a heat pump or a refrigeration system. Although the geometry of the TAHE is simple, the behavior of the engine is complex with 30+ design parameters that affect the performance of the device; therefore, designing such a device remains a significant challenge. In this work, a radical design methodology using reinforcement learning (RL) is employed for the design and optimization of a TAHE for the first time. Reinforcement learning is a machine learning technique that allows optimization by specifying 'good' and 'bad' behavior using a simple reward scheme *r*. Although its framework is simple, it has proved to be a very powerful tool in solving a wide range of complex decision-making/optimization problems. The RL technique employed by the agent in this work is known as Q-learning. Preliminary results have shown the potential of the RL technique to solve this type of complex design problem, as the RL agent was able to figure out the correct configuration of components that would create positive acoustic power output. The learning agent was able to create a design that yielded an acoustic power output of 643.31 W with a thermal efficiency of 3.29%. It is eventually hoped that with increased understanding of the design problem, in terms of the RL framework, it will be possible to ultimately create an autonomous RL agent for the design and optimization of complex TAHEs with minimal predefined conditions/restrictions.

Keywords: thermoacoustic heat engine; heat recovery technology; design; optimization; reinforcement learning

*Corresponding author: harris.makatsoris@brunel.ac.uk

Received 28 November 2014; revised 24 June 2015; accepted 9 July 2015

1 INTRODUCTION

1.1 Background problem

In recent years, there has been a renewed interest in sustainable energy technologies, due to new legislation, continued dependency on fossil fuels and concerns of the negative impact on the environment. One such application of sustainable energy technologies is in manufacturing processes, which result in waste heat, such as chlorine production in the chemical industry, aluminum melting in the materials industry and baking in the food industry [1].

In this particular work, the utilization of waste heat in the food manufacturing baking process is considered. The manufacturing process of biscuit baking results in a certain quantity of gas mixture in the baking oven to be removed, using an extractor fan where it is expelled into the atmosphere in an exhaust gas flue. It is this waste heat from the expelled gas mixture that we are attempting to exploit.

When considering various sustainable energy technologies to design for the application of waste heat recovery in food manufacturing, it is necessary to consider various issues: (1) the geometric limitations of installing such a system in a factory; (2) effective utilization of low-temperature waste heat and (3) the cost of installation and maintenance. As a result of all these considerations, the thermoacoustic heat engine (TAHE) was proposed as a heat recovery technology for this particular application. A TAHE

International Journal of Low-Carbon Technologies 2016, 11, 431–439

doi:10.1093/ijlct/ctv023 Advance Access Publication 16 August 2015

is a type of prime mover that converts thermal power to acoustic power (a type of mechanical power). Unlike other heat recovery technologies for low-temperature waste heat utilization (i.e., Organic Rankine cycle, Kalina cycle), TAHEs have been known to be designed as small as 650 mm in length and 220 mm in height [2], thus providing greater flexibility with regard to where it can be installed into a pre-existing system. Other advantages include 'no moving parts, no exotic materials, and no close tolerances or critical dimensions [3],' making it an attractive heat recovery technology for the food-manufacturing industry, as it will require low initial investment and low maintenance. It can also be used to drive a heat pump, refrigeration system or for electricity generation using a transducer.

The TAHE at its most simplistic is composed of two heat exchangers that is the interface between the heat source and sink and the working fluid, some kind of porous medium where the conversion of energy takes place and a tube that houses the acoustic power produced as shown in Figure 1. The key mechanism for energy conversion from thermal to acoustic is the thermoacoustic effect, occurring in the TAHE when certain conditions are satisfied. A compressible fluid is used as the working fluid within the engine, which in most cases is an inert gas such as helium. Acoustic waves occur naturally as a result of a temperature gradient across the stack as heat transfer occurs between the compressible fluid and a solid boundary (stack). The transfer of thermal energy to and from the compressible fluid and the stack creates local changes of pressure and velocity in the working fluid. When there is the correct pressure–velocity phasing, acoustic oscillations appear spontaneously creating an acoustic wave. Depending on the pressure–velocity phasing, either a standing wave or a traveling wave is created.

Thermoacoustics is an emerging field, and previous research has mostly concentrated on a better understanding of the behavior of such devices, and effective design solutions in an attempt to increase its efficiency. Attempts have been made in the last two decades to realize real-life applications of thermoacoustic devices. One example of such work is the utilization of heat from a four-stroke automobile gasoline engine [4]. Another example is the theoretical work carried out by the Energy Research Centre in Netherlands on a thermoacoustic heat pump for upgrading industrial waste heat [http://www.ecn.nl/fileadmin/ecn/units/eei/Onderzoeksclusters/Restwarmtebenutting/b-07-007.pdf]. Attempts have also been made to design efficient thermoacoustic electricity generators. In these systems, some kind of transducer is coupled with the TAHE to convert acoustic power to electric power. Various designs have explored different methods of transduction, such as piezoelectric [5], magnetohydrodynamic transducers [6] and linear alternators [7] with varying degrees of acoustic-to-electric transduction efficiency (the ratio of electric power output to the acoustic power input) and cost. Thus, previous research conveys the wide range of potential applications of the TAHE.

Although this technology has great potential, there are several main drawbacks that currently hinder the technology from commercialization.

(1) Complex behavior of the TAHE with >30 design parameters that affect the performance of the device has meant that attempts to optimize these numerous design parameters for a specific application remain challenging. For example, for a simple design of a TAHE with parallel plate HXs, a parallel plate stack and a straight tube duct, there are 7 global design parameters (thermal power input, temperature difference across stack, mean operating temperature, mean pressure, peak pressure amplitude, resonant frequency and cross-sectional area of tube), 5 thermophysical design parameters of the working fluid (thermal conductivity, speed of sound, dynamic viscosity, polytropic coefficient and thermal expansion coefficient), 3 thermophysical design parameters each for the stack and the HXs (density, specific

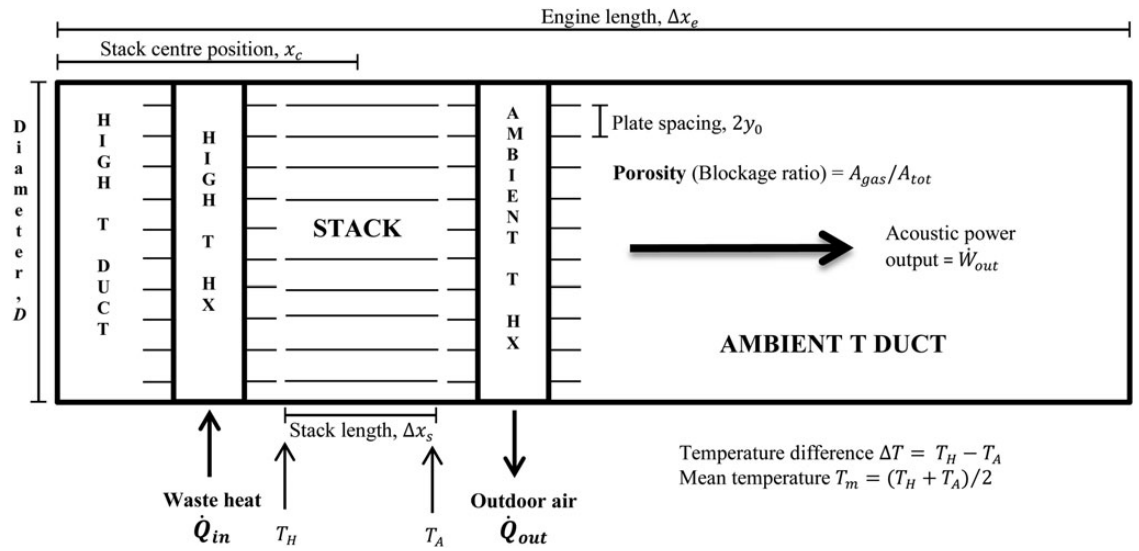

Figure 1. *Design of a TAHE with parallel plate heat exchangers and stack.*

heat and thermal conductivity) and finally 4 geometrical design parameters each for the stack and HXs (plate length, plate center position along *x*-axis, plate spacing and plate thickness); this makes a total of 33 design parameters [8].

(2) Lower efficiencies relative to other more well-established prime movers.

(3) Lack of a complete understanding of the physical behavior of the device.

This work focuses on the first problem, the consideration of the complete design of such a complex energy system, the TAHE, for the utilization of low-temperature waste heat in food manufacturing.

In the next section, previous literature on the design and optimization of TAHEs are presented, their limitations are outlined, and the radical design methodology that is proposed in this paper is introduced. In Section 2, the simulation tool of the TAHE, the design methodology, constraints and implementation are defined. Finally, we will discuss what we have learnt so far from the preliminary study that has been carried out.

1.2 Thermoacoustic heat engine design

Experimental work has yielded greater understanding of the design of the TAHE in order to maximize performance [9–13]. More recently, there have been efforts to design and optimize TAHEs computationally [14, 15]. Wetzel and Herman [8] first proposed a systematic design and optimization algorithm of a standing-wave thermoacoustic refrigerator that provides estimates for initial design calculations. From fundamental equations of linear thermoacoustic theory describing the total power flowing through the stack and the acoustic power produced, it was possible to identify 19 design parameters that affect the performance of the device. These can be categorized into global parameters, material parameters and geometric parameters. This approach was taken further by Babaei and Siddiqui [16], which also takes into account the energy balance equation and entropy balance to improve the optimization process. This type of systematic design methodology was employed for the optimization of a TAHE for the application of low-temperature waste heat recovery in the food manufacturing baking process [17]. Although these works start to consider how to design such a complex energy system in a systematic way, they are still a relatively simplified approach to this complex design problem. These design methodologies start with a predefined design of the standing-wave TAHE and modifies only certain limited aspects of its geometry and thermophysical properties. There are other research that endeavors to optimize the more complex traveling-wave TAHE, such as the work of Ueda *et al.*, focusing on the regenerator aspect of the engine. This work highlights the significant impact of certain key parameters on performance [18]. Also Karimi and Ghobanian attempted the design and optimization of a cascade engine for low-temperature heat sources [19].

Some attempts have been made to employ artificial intelligence optimization techniques in the thermoacoustic field, such as the work described in Srikitsuwan *et al.* and Chaitou and Nika [20, 21], adopting a genetic algorithm and particle swarm optimization method, respectively. Each highlights a few key design parameters that affect the performance of the thermoacoustic device for optimization. While these optimization techniques do not suffer from some of the problems that classical optimization techniques do, they do not scale well at all for problems with increasing levels of complexity, as the size of the search space increases.

This paper describes a radical design methodology using RL. This machine learning technique allows optimization by specifying 'good' and 'bad' behavior using a simple reward scheme *r*, and thereby attaining the desired goal rather than having to explicitly define an objective function like the genetic algorithm. It is able to learn by interacting with an environment. In the RL learning framework, there are two main components: the agent who is the learner and decision maker and an environment that changes state according to the actions taken by the agent. This interaction is depicted in Figure 2.

The goal of the RL agent is to ultimately maximize the accumulation of reward, through a sequence of actions over the long run [23]. Although the learning framework is a simple one, it has proved to be a very powerful tool in solving a wide variety of complex decision-making/optimization problems. For example, it has been used to autonomously learn to play complex games [24] and is also used for control applications [25]. Although RL has not been widely applied to design problems, it is a powerful tool that can handle vast search spaces, optimizing a goal through sampling. Also RL can be used where an analytic solution is not available or where an environment can only be understood by interacting with it, which is why it has been successfully employed for complex problems and why it is used as a design tool in this particular instance.

This work differs greatly from previous research that has attempted to design and optimize TAHEs, because it does not start with a basic design of the TAHE that merely changes certain parameter values. The RL agent must start from knowing nothing of the environment and figure out from its interaction with the modeling tool, the configuration of the device that yields 'good' behavior (i.e., positive acoustic power output). As there is no previous literature regarding the design of an energy system using RL, the focus of this work is to understand how to effectively define the design problem in terms of the RL framework, which in itself greatly affects the outcome. As a benchmark for the results obtained using RL, these will be compared with the results described in Mumith *et al.* [17].

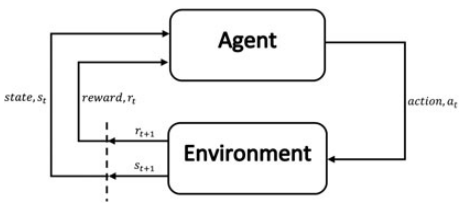

Figure 2. *The agent–environment interaction in RL [22].*

2 METHODS

2.1 Simulation of environment

The environment in this particular RL problem is DeltaEC [26], a simulation code that provides information regarding the performance of thermoacoustic equipment. This modeling tool has been adopted because it is widely used in research to aid in the design of thermoacoustic devices, to achieve desired performance. It also allows simple actions to be made in order to modify the design, due to the fact that it is represented by a series of segments, such as duct, stack and heat exchanger.

The numerical integration method that is employed to calculate output values integrates in one-spatial dimension using a low-amplitude, 'acoustic' approximation and sinusoidal dependence [26]. The wave equation without viscous or thermal-hysteresis losses is as follows:

$$p_1 + \frac{a^2}{\omega^2}\frac{\mathrm{d}^2 p_1}{\mathrm{d}x^2} = 0 \tag{1}$$

This second-order equation can be considered as two first-order equations with respect to pressure p_1 and volume flow rate U_1:

$$\frac{\mathrm{d}\, p_1}{\mathrm{d}x} = -\frac{i\omega\rho_\mathrm{m}}{A} U_1 \tag{2}$$

$$\frac{\mathrm{d}U_1}{\mathrm{d}x} = -\frac{i\omega A}{\rho_\mathrm{m} a^2} p_1 \tag{3}$$

These are the most basic and fundamental equations required to find the pressure and volume flow rate of the system as a function of x. DeltaEC uses more complex equations based on Equations 2 and 3, which account for acoustic power dissipation in ducts, and uses varying equations for different segments to suit local conditions.

Acoustic power (also known as the work flux) generated in a TAHE is related to the work done by a differential volume of fluid $\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ as it expands from $\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ to $\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z + \mathrm{d}V$, and the work is pdV. The acoustic power is simply the time-averaged product of p_1 and U_1 as shown below:

$$\dot{E} = \frac{1}{2}\mathrm{Re}[p_1 \tilde{U}_1] \tag{4}$$

Equations 2 and 3 are integrated and are used to calculate the acoustic power according to Equation 4, at a particular location in the engine along the x-axis. As Equations 2 and 3 depend on the type of segment, the acoustic power produced at a particular location in the engine also depends on the type of segment that is being described at that location.

The segments that can be chosen by the agent during the design of the TAHE are those that are also used in Mumith *et al.* [17], the DUCT segment is the geometric container that houses the acoustic wave, the STKSLAB segment is a stack composed of parallel plates and the HX segment is a heat exchanger, also composed of parallel plates.

2.2 The RL problem

2.2.1 *States, actions and rewards*

The interaction between the environment and agent occurs in discrete time steps known as episodes (can also be described as an interaction). At each episode, the environment starts in state s_t, the agent takes an action a_t, a reward r_{t+1} is given and the state of the environment changes to s_{t+1}. How the states and actions are represented in terms of the RL framework greatly affects the success of the solution produced by the agent and the computational complexity of the problem. Therefore, these have been kept as simple as possible in this preliminary study.

The states are based on the work described in Mumith *et al.* [17], which was able to achieve ~1 kW of acoustic power output after optimization. The acoustic power output is divided into three possible states. First, the undesirable state (=Negative), where negative power output occurs (when the engine absorbs acoustic power), then the Low state where positive acoustic power is created but less than 1 kW and finally the desirable state (=High), which is what we are really interested in. The conditions for each state are summarized below:

State, S = {high, low, negative}

High if $\dot{W}_\mathrm{out} > 1\,\mathrm{kW}$ acoustic power output (for $\dot{Q}_\mathrm{in} = 19\,\mathrm{kW}$ and $T_\mathrm{A} = 278\,\mathrm{K}$)

Low if $0 < \dot{W}_\mathrm{out} \leq 1\,\mathrm{kW}$ for 19 kW thermal power input (for $\dot{Q}_\mathrm{in} = 19\,\mathrm{kW}$ and $T_\mathrm{A} = 278\,\mathrm{K}$) Negative if $\dot{W}_\mathrm{out} < 0$ for 19 kW thermal power input (for $\dot{Q}_\mathrm{in} = 19$ kW and $T_\mathrm{A} = 278\,\mathrm{K}$)

As a result of the defined state s_t, the agent makes a decision of what the next action is depending on how it learns and explores the search space. The actions that can be taken resemble the design process that is carried out in the simulation tool DeltaEC.

Action, A = {delete segment, add segment, increase parameter value, decrease parameter value}

When the reward is allocated and the magnitude of the reward will have a significant effect on what the RL agent learns. Therefore, we have chosen a reward scheme that we believe will encourage the agent, to create a design that surpasses the performance of the TAHE that is described in Mumith *et al.* [17]. We have assigned the greatest reward if it can achieve an acoustic power output greater than 1 kW. In RL, negative values can be assigned to discourage unwanted behavior; therefore, a reward of −50 has been chosen whenever the RL agent observes a negative acoustic power output value. A relatively low reward is assigned whenever the environment is at a Low state $\dot{W}_\mathrm{out} \leq 1\,\mathrm{kW}$. This is another way to implicitly encourage the agent to choose actions that can yield $\dot{W}_\mathrm{out} > 1\,\mathrm{kW}$. Naturally, the greatest reward is assigned to the desirable state, High, when $\dot{W}_\mathrm{out} > 1\,\mathrm{kW}$. The magnitude of the rewards assigned to the each state represents the level of desirability to be in that particular state.

Reward, R (high) = +50

Reward, R (low) = +5

Reward, R (negative) = −50

The history of the interaction between the environment and agent can be considered a sequence of state-action rewards:

$$(s_t,\ a_t,\ r_{t+1},\ s_{t+1},\ a_{t+1},\ r_{t+2},\ s_{t+2},\ a_{t+2}\ \ldots).$$

For each episode, we have the information:

$$(s_t a_t,\ r_{t+1},\ \ s_{t+1}).$$

Therefore, the agent learns from its historical interaction with the environment and so ultimately attempts to maximize the sum of expected rewards.

$$R_t = r_{t+1} + \gamma^1 r_{t+2} + \gamma^2 r_{t+3} + \ldots + \gamma^k r_{t+k+1} \qquad (5)$$

where k is the number of interactions in which a reward was received from the environment to the agent, and γ is the discount rate that provides weighting to future expected rewards ($0 \leq \gamma \leq 1$). If γ is 0, then the agent's objective is to only maximize the immediate reward, but as γ increases toward the value 1, the objective is shifted to consider future rewards more strongly as the agent becomes more farsighted. Finally, when $\gamma = 1$ all future rewards are considered equally [22, 23].

2.2.2 *Q-Learning*

The way in which the RL agent learns and explores the environment (i.e., DeltaEC) is at the heart of this machine learning technique. Most RL algorithms are based on estimating value functions, which estimates how good a decision is (i.e., which action to take) depending on the current state of the environment, in terms of expected rewards. For each state-action pair, an estimated action-value function $Q(s_t, a_t)$ is calculated and updated every time $s_t = s$ and $a_t = a$. In this particular problem, there are 18 possible actions that can be taken by the RL agent (3 segments can be added/deleted and 6 parameter values can be increased/decreased) and 3 states. Therefore, there are altogether 54 state-action pairs for which the estimated action-value function must be calculated. In this current work, these data are stored and retrieved from a lookup table, but this very quickly becomes impractical as the number of states and actions increases.

The value function learning technique used in this work is known as Q-Learning, which directly approximates the optimal action-value function. It is an RL technique that can learn by simply sampling the state space and does not require a complete probability distribution over the actions to all states.

$$Q\,(s_t,\ a_t) \leftarrow Q\,(s_t,\ a_t) + \alpha\,[r_{t+1} + \gamma\ \max_a(s_{t+1},\ a) - Q\,(s_t,\ a_t)] \qquad (6)$$

where $\alpha (0 < \alpha \leq 1)$ is the learning rate (also known as step size), which determines the weighting given to newly acquired information compared with previous information, and thus allows the RL agent to account for changes in the nonstationary environment (i.e., the DeltaEC model) as time passes [24]. The learning rate and discount rate are two main parameters that determine how the RL agent learns and makes decisions. A summary of the Q-learning algorithm is shown in Figure 3.

The agent must not only exploit the latest information by choosing the next action based on the optimal action-value function $Q^*(s,\ a) = \max\ Q(s,\ a)$ but must also explore the search space of potential TAHE designs and parameter values, for actions that could potentially result in long-term future rewards (i.e., maximal performance of the TAHE). A simple way to do this is to adopt an ε-greedy method, which uses a simple tuning parameter $0 \leq \varepsilon \leq 1$, where ε is the probability that a random action will be chosen and $1 - \varepsilon$ is the probability that a greedy action, i.e., one that maximizes $Q(s, a)$, will be chosen at a particular time [22].

The conditions of the RL problem have been set as follows:

(1) Discount rate, $\gamma = 0.99$.
(2) Learning rate, $\alpha = 0.1$.
(3) Exploration parameter, $\varepsilon = 0, 0.2$.
(4) All $Q(s, a)$ values are initialized at 0.
(5) Equiprobable random policy employed (all actions are equally likely).

2.3 Design constraints

Unlike when an RL agent interacts for example with a game, which tells it when it makes an illegal move, in this case the environment (DeltaEC) cannot tell the agent if an illegal move is carried out (i.e., physically impossible parameter values). Hence, parameter values are restricted beforehand to a range of values that allow physically meaningful results, shown in Table 1. As this is a comparison between the works described in Mumith *et al.* [17], certain parameter values are kept the same. The range of values shown in Table 1 is based on previous research [8]. If an action chosen by RL agent would result in the model going beyond the design constraints, then another action is chosen.

Any TAHE requires at the very least four components in order to function; a stack, two heat exchangers and a solid container.

Initialize Q (s,a) arbitrarily
Repeat (for each episode):
    Initialize s
    Repeat (for each step of episode):
        Choose a from s using policy derived from Q (e.g. ϵ -greedy)
        Take action a, observe r, s'
        $Q(s,a) \leftarrow Q(s,a) + \alpha\,[r + \gamma\, Max_{a'}(s', a') - Q(s,a)\,]$
        $s \leftarrow s'$;
    Until s is terminal

Figure 3. *Q-learning algorithm [13].*

Table 1. *Design parameters*

	Design parameters	Value/range	±
BEGIN segment			
1	Mean pressure Mean P, P_m (MPa)	1.00–3.00	0.1
2	Resonant frequency Freq, f_r (Hz)	$a/2\Delta x_e$ where a is the speed of sound (m/s) and Δx_e is the total length of the engine (m) 43.38–548.25	Calculated
3	Mean temperature Beg, T_m (K)	423	n/a
4	Pressure amplitude $\|p\|$, $\|p_1\|$ (Pa)	DR × P_m 0.005–0.06	0.005
5	Phase of pressure amplitude Ph(p), (°)	90	n/a
6	Velocity amplitude $\|U\|$, $\|U_1\|$ (m^2/s)	0	n/a
7	Phase of velocity amplitude Ph(U), (°)	0	n/a
8	Gas (type)	Helium–argon mixture	n/a
9	nL, X_A	0.0–1.0	0.05
SURFACE segment			
10	Cross-sectional area, A (m^2)	0.005	n/a
DUCT segment			
11	Cross-sectional area, A (m^2)	0.005	n/a
12	Perimeter, Π (m)	0.194–0.614	Calculated according to cross-sectional area
13	Length, Δx_{duct} (m)	0.50–3.00	0.1
14	Surface roughness	5×10^{-4}	n/a
STKSLAB segment			
15	Total cross-sectional area	0.005	n/a
16	Porosity (Blockage ratio, BR_{stack})	0.8	n/a
17	Length, Δx_{stack} (m)	0.10–0.25	0.01
18	Half plates spacing, y_0 (m)	δ_κ	n/a
19	Half thickness of solid plate, l (m)	$L_{plate} = y_0(1 - BR)/BR = y_0/4$	Calculated
20	Plate material, $Solid_{stack}$	Stainless steel	n/a
HX segment			
21	Total cross-sectional area	0.005	n/a
22	Porosity (Blockage ratio, BR_{HX})	0.4	n/a
23	Length, Δx_{HX} (m)	0.02–0.06	0.005
24	Half plates spacing, y_0 (m)	δ_κ	n/a
25	HeatIn, $\dot{Q}_{in}$ (kW) (Thermal power input)	19.0	n/a
26	HeatIn, $\dot{Q}_{out}$ (kW) (Thermal power output)	Set as guess in DeltaEC	n/a
27	Solid T, T_A (K) (Temperature of plates of ambient heat exchanger)	278 Set as target	n/a
28	Solid material, $Solid_{HX}$	Copper	n/a

Therefore, a restriction of a minimum of four segments is imposed on the model. Also the maximum number of segments allowed was set at 8, to ensure that the model does not exhibit wildly unrealistic physical behavior. When state is initialized, four segments are chosen at random in no particular order. The parameter values are also chosen at random within the range of values specified in Table 1. All DeltaEC models require specific segments: the BEGIN segment, which sets global parameters, the HARDEND segment that denotes the beginning or end of a solid container and the SURFACE segment, which comes before the HARDEND segment and accounts for thermal-hysteresis dissipation [26]. These are not included in the restrictions of the number of segments in a model.

2.4 Implementation

The RL design problem has been modularized as shown in Figure 4, each with its own distinct job.

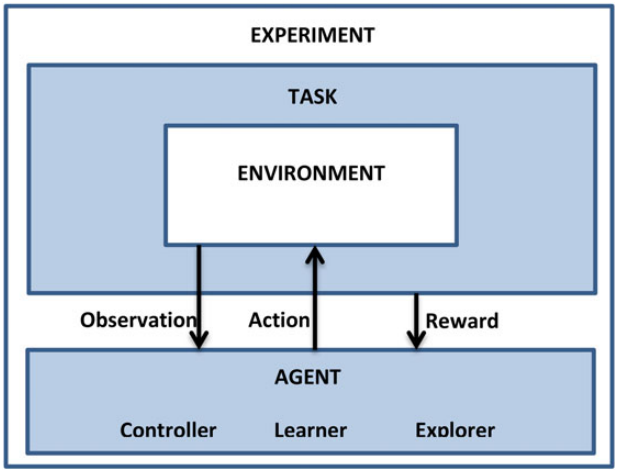

Figure 4. *Various elements of the RL problem.*

- Environment (DeltaEC): this is the world with which the agent interacts. It also defines the constraints imposed on the world that are specified in the previous section.
- Task: it handles the interaction between the agent and environment, defines the ultimate goal of the environment (i.e., outlines the reward scheme) and decides when an episode is over.
- Agent: the agent has its own learning component (Q-learning), a controller that stores and retrieves Q-values from a lookup table, and an explorative component (ε-greedy method).
- Experiment: the experiment brings the environment, task and agent together so that they interact as shown in Figure 4 to create the RL problem.

The actions chosen by the RL agent is executed using the Python toolkit WATSUP [27], which directly interacts with the DeltaEC window application to find and invoke actions on controls and menu items.

3 RESULTS AND DISCUSSION

In RL, the average reward (sum of total rewards/total number of episodes) gives a general indication of the relative success of the RL agent to solve a problem, the higher the average reward the better the RL agent is deemed to have performed. In this preliminary study, the number of episodes carried out was 300 when $\varepsilon = 0, 0.2$. After just 50 episodes, the average reward obtained by the RL agent when $\varepsilon = 0$ is -37.9, whereas the average reward when $\varepsilon = 0.2$ is -12.6. The greatest acoustic power output that the greedy policy ($\varepsilon = 0$) was able to achieve from a DeltaEC model was 69.38 W, whereas the ε-greedy policy ($\varepsilon = 0.2$) yielded a maximum of 643.31 W. In both cases, the thermal power input of the engine that is designed is 19 kW and the temperature difference across the stack is 145 K ($\Delta T = 423$ K $-$ 278 K $=$ 145 K). Therefore, the thermal efficiency of the TAHE that is designed in DeltaEC when the ε-greedy policy is employed is 3.29% and is $\sim$10% of the Carnot efficiency, whereas the thermal efficiency is 0.37% and $\sim$1% of the Carnot efficiency when the greedy policy is employed. If we make a direct comparison with results obtained from Mumith *et al.* [17], while the use of the RL technique does not currently outperform the iterative design methodology, there are promising signs for the future application of RL for the design and optimization of energy systems, as the agent was able to correctly determine the configuration of segments in DeltaEC that would yield positive acoustic power output, through the continual feedback after each action of its corresponding reward. Also it is believed that greater interactions and time exploring the environment (i.e., more episodes) in future work will allow the agent to better understand the DeltaEC environment that it is interacting with, and therefore be able to better understand the complexities of designing the TAHE.

The greedy agent ($\varepsilon = 0$) performed poorly as it would choose actions that had optimal action-value functions, even though these values were estimates and did not necessarily reflect their true values at the time of the interaction between DeltaEC and the RL agent. So for example if an action yielded a good reward, then there was a propensity for the agent to keep choosing the same action, particularly for the 100 or so episodes, as the RL agent believed that it would yield optimal results for the next interaction with DeltaEC based on the maximum q-value for a particular state-action pair. These results show that it is imperative for any search algorithm to thoroughly explore the search space in order to evaluate wildly different design options, so that it can ultimately choose an optimal design choice. The problem with the ε-greedy method was that if it chose an action at random, then all actions were equally likely to be chosen; therefore, it is just as likely to choose a bad action as a good one. Even though this exploration approach allowed the RL agent to learn more about how various actions affected the performance of the TAHE, in the short term, it resulted in certain designs that yielded unsatisfactory performance. It is difficult to comprehend the extent of the problem until greater numbers of episodes are carried out. It may be that it is not such an issue with more interactions between the environment and agent. But it may be necessary to adopt a more sophisticated exploration technique such as the softmax method, which determines the probability of selecting an action based on its q-value (the higher the q-value the greater the probability that an action is chosen), and therefore an equiprobable random policy is not employed. Also another issue is the lack of constraints with regard to ordering of the segments in DeltaEC. Again, as with the previous problem, it is difficult to determine how much of a problem this will be when greater number of episodes are carried out, otherwise it may be necessary to employ sequencing constraints, so that the RL agent does not merely produce models that are physically meaningless. In both of the experiments, the high state was never observed; this may be partly to do with the number of episodes carried out during the experiments, but this could also be because the condition for the environment to be in a High state may be unrealistically high. Even though great caution was taken so that the DeltaEC model would produce physically meaningful results, at times it was difficult to match the temperatures of the heat exchangers to desired values due to the random way in which

the segments were assembled, and because the upper limit of the stack length was too high.

The results show that the structure of the design problem in terms of the RL framework was generally successful, as the RL agent was able to learn the correct configuration of a TAHE in order to produce positive acoustic power output. Although >300 episodes must be carried out to reach a definitive conclusion, we believe that we have not yet exploited the full and expansive capabilities of the RL technique. The behavior of the RL agent itself must be optimized for this particular application, which means further comprehensive study tuning the various learning and exploration parameters of the RL agent. Also the way in which the design problem was defined in terms of the RL framework must be amended to be able to manage more adequately the complexities of the design problem. This can only be understood by experimenting with the reward scheme, definitions of state, etc. For example, reward can be given according to the magnitude by which the acoustic power increases/decreases with respect to the previous episode, rather than the state with which it is in at the time. Also, when a greater number of episodes is carried out, it will be possible to reduce the steps by which the parameter values are increased or decreased, so that it is easier for the RL agent to track changes to the environment as a result of actions taken, and learn more effectively how these changes affect the performance of the TAHE.

4 CONCLUSIONS

The TAHE is an energy technology that can potentially be very useful in a wide range of applications, but currently its complex physical behavior and the many design parameters that affect its performance means that designing such a complex energy system is challenging. In previous research, systematic design algorithms can only be employed in limited circumstances with a basic design already predefined. Therefore, this paper outlines for the first time the implementation of a radical design methodology using RL, where the RL agent learns by itself from its interaction with the DeltaEC model, good and bad behavior in terms of acoustic power output, in order to design and optimize a TAHE from scratch.

Preliminary results have shown the potential of the RL technique to solve this type of complex design problem, as the RL agent was able to figure out the correct configuration of components that would create positive acoustic power output. The learning agent was able to create a design that yielded an acoustic power output of 643.31 W, with a thermal efficiency of 3.29%, when the temperature difference across the stack is 145 K.

It is necessary to experiment further with various aspects of the way the design problem has been defined in terms of the RL framework and the learning and exploration parameters of the RL agent, in order to fully understand how certain parameters affect what and how the RL agent learns. It is hoped that with increased understanding, we can eventually achieve our ultimate goal, which is an autonomous agent for the design and optimization of more complex TAHEs, such as the traveling-wave TAHE, which is inherently more efficient, but more complicated to design and is currently the real attraction of TAHE research. We believe that the ability of RL to effectively search vast search spaces and to evaluate a large number of configurations and designs of the TAHE lends itself to ultimately be able come up with a novel design.

REFERENCES

[1] U.S. Department of energy, 2005. Opportunity analysis for recovering energy from industrial waste heat and emissions. [online] www.eere.energy.gov/industry/imf/pdfs/4_industrialwasteheat.pdf (20 June 2010, date last accessed).

[2] Zhou G, Li Q, Li ZY, *et al*. A miniature thermoacoustic stirling engine. *Energy Convers Manag* 2008;**49**:1785–92.

[3] Swift GW. Thermoacoustic engines. *J Acoust Soc Am* 1988;**84**:1145–80.

[4] Hatazawa M, Sugita H, Ogawa T, *et al*. Performance of a thermoacoustic sound wave generator driven with waste heat of automobile gasoline engine. *Trans Jpn Soc Mech Eng* 2004;**70**:292–9.

[5] Jenson C, Raspet R. Thermoacoustic power conversion using a piezoelectric transducer. *J Acoust Soc Am* 2010;**128**:98–103.

[6] U.S Patent No. 4,559,551.

[7] Backhaus S. Traveling-wave thermoacoustic electric generator. *Appli Phys* 2004;**85**:1085–7.

[8] Wetzel M, Herman C. Design optimization of thermoacoustic refrigerators. *Int J Refrig* 1997;**20**:3–21.

[9] Swift GW. Analysis and performance of a large thermoacoustic engine. *J Acoust Soc Am* 1992;**92**:1551–63.

[10] Hariharan NM, Sivashanmugam P, Kasthurirengan S. Optimization of thermoacoustic primemover using response surface methodology. *HVAC&R Res* 2012;**18**:890–903.

[11] Backhaus S, Swift GW. A thermoacoustic Stirling heat engine. *Nature* 1999;**399**:335–8.

[12] Tijani MEH, Spoelstra S. A high performance thermoacoustic engine. *J Appl Phys* 2011;**110**:093519.

[13] Abduljalil AS, Yu Z, Jaworski AJ. Design and experimental validation of looped-tube thermoacoustic engine. *J Therm Sci* 2011;**20**:423–9.

[14] Trapp AC, Zink F, Prokopyev OA, *et al*. Thermoacoustic heat engine modeling and design optimization. *Appl Therm Eng* 2011;**31**:2518–28.

[15] Tartibu LK, Sun B, Kaunda MAE. Multi-objective optimization of the stack of a thermoacoustic engine using GAMS. *Appl Soft Comput* 2015;**28**:30–43.

[16] Babaei H, Siddiqui K. Design and optimization of thermoacoustic devices. *Energy Convers Manag* 2008;**49**:3585–98.

[17] Mumith JA, Makatsoris C, Karayiannis TG. Design of a thermoacoustic heat engine for low temperature waste heat recovery in food manufacturing. *Appl Therm Eng* 2014;**65**:588–96.

[18] Ueda Y, Mehdi BM, Tsuji K, *et al*. Optimization of the regenerator of a traveling-wave thermoacoustic refrigerator. *J Appl Phys* 2010;**107**:034901.

[19] Karimi M, Ghorbanian K. Design and optimization of a cascade thermoacoustic Stirling engine. *Proc Inst Mech Eng A J Power Energy* 2013;**227**:814–24.

[20] Srikitsuwan S, Kuntanapreeda S, Vallikul P. A genetic algorithm for optimization design of thermoacoustic refrigerators. In *Proceedings of 7th WSEAS International Conference on Simulation, Modelling and Optimization*, Beijing, China, 2007, 15–7.

[21] Chaitou H, Nika P. Exergetic optimization of a thermoacoustic engine using the particle swarm optimization method. *Energy Convers Manag* 2012;**55**:71–80.
[22] Sutton RS, Barto AG. *Reinforcement Learning: An Introduction*. Vol. **1**. No. 1. Cambridge: MIT pressCambridge, 1998.
[23] Woergoetter F, Porr B. *Reinforcement learning. Scholarpedia* 2008;**3**:1448.
[24] Veness J, Ng KS, Hutter M, *et al*. A Monte-Carlo aixi approximation. *J Artif Intell Res* 2011;**40**:95–142.
[25] Hafner R, Riedmiller M. Reinforcement learning in feedback control. *Mach Learn* 2011;**84**:137–69.
[26] Ward B, Clark J, Swift G. 2008. Design environment for low-amplitude thermoacoustic energy conversion: DeltaEC version 6.2 User Guide. [e-book] http://www.lanl.gov/thermoacoustics/DeltaEC.html (July 2010, date last accessed).
[27] https://groups.google.com/forum/#!topic/riverbird2005/p6KspI_u_Vs (16 June 2014, date last accessed).